

Section 4: Population Scale Analysis

One sample is obviously not enough to know what is happening in a population. You are interested in assessing genetic differences on a population scale.

So, you processed about ~230 samples and did the normalization on a genome level. Now, you want to find whether there is any association of the 4 asthma-associated SNPs (rs8067378...) on ORMDL3 expression.

Q13: Read this file into R and determine the sample size for each genotype and their corresponding median expression levels for each of these genotypes.

```
expr <- read.table("rs8067378_ENSG00000172057.6.txt")
head(expr)
```

```
##      sample geno      exp
## 1 HG00367   A/G 28.96038
## 2 NA20768   A/G 20.24449
## 3 HG00361   A/A 31.32628
## 4 HG00135   A/A 34.11169
## 5 NA18870   G/G 18.25141
## 6 NA11993   A/A 32.89721
```

```
nrow(expr)
```

```
## [1] 462
```

The total sample size of the dataset is 462.

```
table(expr$geno)
```

```
##
## A/A A/G G/G
## 108 233 121
```

A/A genotype has the sample size of 108. A/G genotype has the sample size of 233. G/G genotype has the sample size of 121.

```
summary(expr$exp)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   6.675  20.004   25.116   25.640   30.779   51.518
```

```
aggregate(exp ~ geno, data = expr, FUN = median)
```

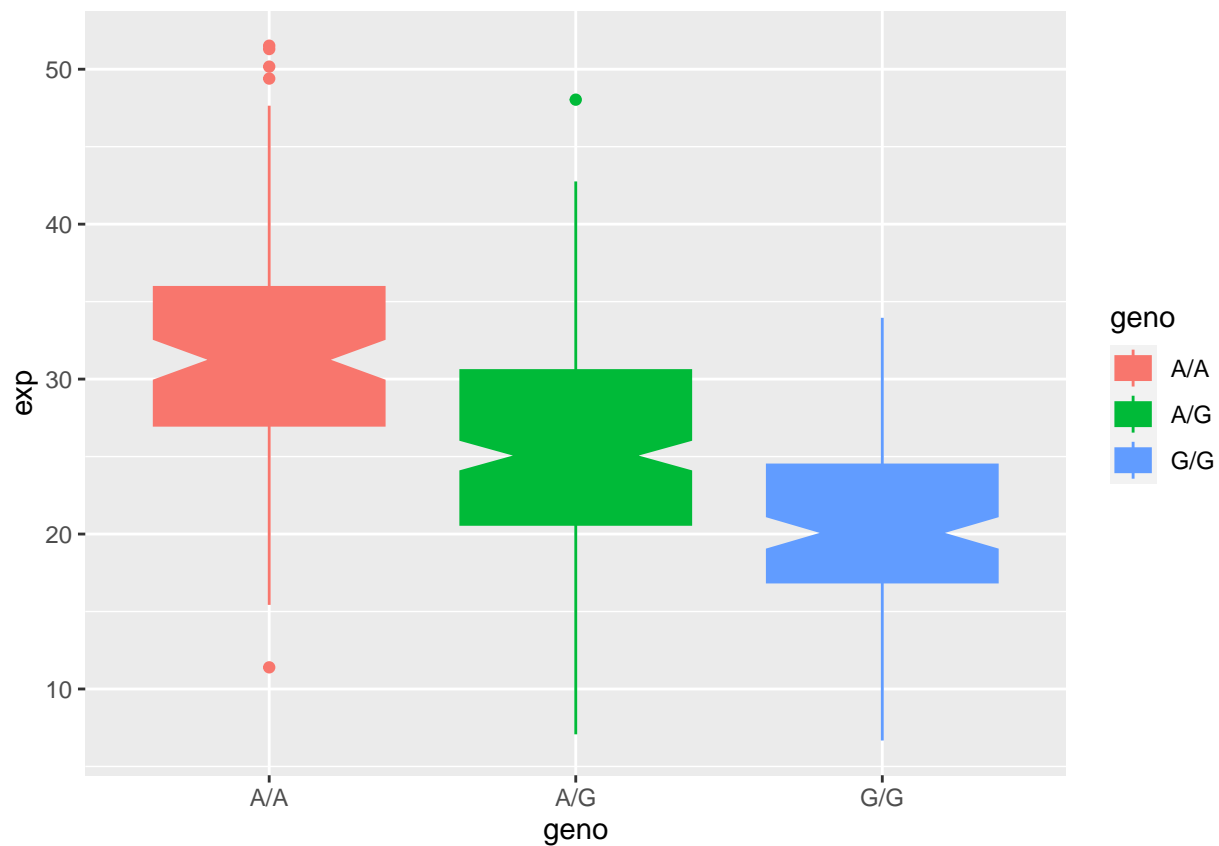
```
##      geno      exp
## 1   A/A 31.24847
## 2   A/G 25.06486
## 3   G/G 20.07363
```

Median expression level of the A/A genotype is 31.25. Median expression level of the A/G genotype is 25.06
Median expression level of the G/G genotype is 20.07

Q14: Generate a boxplot with a box per genotype, what could you infer from the relative expression value between A/A and G/G displayed in this plot? Does the SNP effect the expression of ORMDL3?

```
library(ggplot2)
```

```
ggplot(expr) + aes(geno, exp, col=geno, fill=geno) +  
  geom_boxplot(notch=TRUE)
```



The boxplot shows that expression is a lot higher with A/A genotype compared to G/G genotype. Thus, the SNP affects the expression of ORMDL3.